BSES Limited

FINAL REPORT - SRDC PROJECT STU060
ESTIMATES OF BREEDING VALUE OF SUGARCANE CLONES AND THEIR IMPACT
ON EFFICIENT PARENT MANAGEMENT AND CROSS POLLINATION
by
FELICITY ATKIN
SD10002

Contact:
Felicity Atkin
Research Officer
BSES Limited
PO Box 86
Indooroopilly Q 4068
Telephone: 07 3331 3319
Facsimile: 07 3871 0383
Email: fatkin@bses.org.au

CONTENTS

SUMMARY

Sugarcane breeders in Australia use family trials from four regional programs to estimate breeding values of current parents. Genetic gain to the Australian sugarcane industry can be maximised though efficient parent selection and choosing specific family combinations for cross pollination.

Previously, breeding values were estimated for each region independently using a statistical approach developed over 10 years ago based on a selection index called Net Merit Grade. A new statistical approach was developed to improve breeding value estimates by addressing many of the limitations of the previous approach, including: accounting for site-specific spatial variation; combining family trials across regions using a Factor-Analytic mixed model to exploit genetic correlations between regions; and estimate breeding values for individual traits, cane yield and sucrose content, rather than Net Merit Grade. The new breeding values were included in an economic selection index developed by sugarcane breeders to replace Net Merit Grade, since Net Merit Grade no longer reflected the sugarcane production systems.

Simulated selection of parents using the new statistical approach suggest that additional increases in genetic gains of up to 29% for cane yield and 10% for sucrose content could be achieved when parents are selected on breeding values estimated using the new approach over the previous approach. Additional gains could be achieved if parents are selected using the new economic selection index combined with breeding values estimated using the new approach. Sugarcane breeders could also select parents with greater confidence when using the new statistical approach.

This project has resulted in a new system of estimating the breeding potential of sugarcane parents in the BSES-CSIRO breeding program. More accurate assessment and selection of sugarcane parents for cross pollination is likely to result in increased genetic gain through managing parent populations more efficiently, selecting better cross combinations, and advancement of superior varieties into later stages of selection.

The project has formed the bases of a PhD thesis submitted to The University of Queensland.

1.0 BACKGROUND

Identifying the best parents for breeding is fundamental to achieving genetic gain in any breeding program. As such, optimising the prediction of breeding values (BV) of sugarcane parents is a primary objective of the BSES-CSIRO Plant Improvement program. BSES operates four regional programs to develop new sugarcane varieties for the Australian sugarcane industry. Family evaluation occurs in the first stage of a three-stage selection system in family trials, Progeny Assessment Trials (PATs), to assess the breeding potential of current sugarcane parents. BVs estimated from PATs are used by BSES plant breeders to optimise the gain and efficiency of population improvement. This is achieved through efficient selection of superior parents, choosing specific cross combinations, and allocating resources efficiently to maximise the crosses made using the best parents available. This is one of the key components to the success of the BSES breeding program.

One important recommendation arising from the project "Evaluation and re-structuring of regional selection programs to maximise efficiency and speed of cultivar release" (SRDC project CTA028) was "to conduct future research to examine the value of retaining separate regionally based parent populations and crossing operations, and the relative value of breeding for specific regional adaptation versus broad adaptation" (Recommendation 3). Only clone by environment interactions were investigated in CTA028. However, there was no examination of family x environment (or breeding value x region) interactions and this is the focus of my research.

In this project, a new statistical approach was developed by addressing many of the issues identified with the previous statistical approach to estimate BVs of sugarcane parents from family data by using a Factor-Analytic (FA) mixed model (Kelly et al. 2009; Smith et al. 2001) while modelling spatial variation (Gilmour et al. 1997; Stefanova et al. 2009), to combine family data across regions and allow estimation of separate genetic variances in each region and different genetic correlations between regions.

2.0 OBJECTIVES

This project aimed to improve precision in sugarcane parent selection for cross pollination through improving the accuracy of parental BV predictions.

The specific objectives of the project were to:
- Assess the nature and magnitude of family (or BV) by region interaction in stage 1 trials in the sugarcane breeding program;
- Determine the minimum number of regions required to maximise genetic gain;
- Identify opportunities to link stage 1 selection sites between regions by determining whether similar family (or parent) selections are made at different regions

All of the objectives of the project have been achieved.

Objective 1 - Assess the nature and magnitude of family (or BV)-by-region interaction in stage 1 trials in the sugarcane breeding program.

By using a Factor-Analytic mixed model (Kelly et al. 2009; Smith et al. 2001), by allowing each region to have a different genetic variance and different genetic correlations between pairs of regions, I was able to estimate the magnitude of genotype-by-environment (GxE) interactions for BVs between regions for both cane yield (tonnes cane per hectare, TCH) and sucrose content (commercial cane sugar, CCS). GxE between regions were of similar size to GxE within regions. This suggests that the major factors driving GxE are not region specific,

but rather, are sampled within any given region. These results are similar to the results of CTA028 where they examined GxE interaction of sugarcane clones across regions.

Objective 2 - Determine the minimum number of regions required to maximise genetic gain.

This is addressed in Objective 3.

Objective 3 - Identify opportunities to link stage 1 selection sites between regions by determining whether similar family (or parent) selections are made at different regions.

Genetic correlations between pairs of regions were estimated while assessing GxE between regions. These genetic correlations between regions for both TCH and CCS were high. In addition, the statistical method used in this project allowed me to predict BVs for each parent in each region and make direct comparisons by simulating parent selection. A good proportion of top performing parents were in common for both TCH and CCS when parents were selected within each regional program; however, BVs predicted in the Burdekin region seemed to be the least similar to the other regions. Even though there were some differences in parent selection between regions, there is still the potential to either indirectly select parents from other regions, or to combine some or all regional parent populations and amalgamate all crossing procedures by selecting parents for the industry as a whole. There is also the possibility that resources used in planting family trials could be shared across regions, including either growing common families across regions, or combining family trials across regions. However, issues such as quarantine restrictions within the industry also need to be considered in addition to genetic correlations between regions when deciding which regions are best suited to combining.

3.0   RESEARCH METHODOLOGY, RESULTS AND DISCUSSION

The following summarises material to be presented in my PhD thesis. A copy of the thesis will be sent to SRDC on completion of the degree. A draft of the thesis is contained in Appendix 1 (electronic version only).

3.1.   Efficient use of sugarcane family and pedigree information

3.1.1. Materials and methods

Ten years of sugarcane family trials (i.e. Progeny Assessment Trials, PATs) from the Southern BSES selection program in Bundaberg were used, from 1996 to 2005. A total of 241,520 seedlings from 2318 families were grown in 20 trials over this 10-year period. The number of families grown each year varied from 233 to 411 full-sib families and were a mixture of proven (i.e. progeny had performed well in the past) and experimental (or untested) combinations. The number of families in common across years varied between 0% and 27% as a result of sparse and unreliable flowering and poor pollen fertility in sugarcane breeding (Berding and Skinner 1987), but there was considerably greater connectivity across years amongst the parents (15% to 62%).

Experimental data available for each of the 20 trials at time of harvest included sucrose content (CCS) and cane yield (TCH) of each family plot. However, only data on CCS was considered here as interplot competition is known to substantially affect the estimation of TCH in sugarcane (Stringer and Cullis 2002), and a method to model competition when combining sugarcane family data across trials has not been fully developed.

Ten data sets were generated, each representing different numbers of years available. The first data set contained only the 2005 trials (i.e. the most recent trial data). The second data set contained 2 years of data, i.e. 2005 and 2004. Each subsequent data set included the next year of trials, so that the last data set contained data from all 10 years of trials planted from 1996 to 2005. Pedigree information available for each sugarcane parent back to their founding populations (Roach 1989) was also used to generate pedigree files for each of the 10 data sets. Six pedigree files were generated for each data set, each containing a different depth of pedigree. The first pedigree contained only the parents of each family grown; labelled two (2) generations (i.e. of the seedlings and their parents). The second pedigree generated for each data set included an additional generation of parents (i.e. the grandparents of each family). This pedigree level was labelled three (3) generations. For each depth of pedigree, an additional generation of parents was added, until there were seven generations included in the pedigree. A pedigree of seven additional generations was considered a complete pedigree as inclusions of pedigree information beyond seven generations back to the founding parents only resulted in the addition of two new parents.

3.1.2. Statistical analysis

A modelling process presented in Gilmour et al. (1997) was also applied to CCS to include site-specific modelling terms due to spatial variation in each trial. A genetic model, the parental model (Henderson 1984; Mrode 2005), was used to partition the random genotypic effects of unique parents, as a sugarcane parent can be used as either a male or a female, or both. Genetic effects of each parent were further partitioned into additive and non-additive genetic effects as per Costa e Silva et al. (2004).

For each of the 10 data sets, CCS were analysed using the parental model. As some parents act as both male and female in sugarcane, a unified estimate for each parent was obtained. Each of the 10 data sets was analysed using the six different depths of pedigree for each of these two models, with a total of 120 analyses performed. All analyses were performed using ASReml (Gilmour et al. 2006).

Additive genetic components estimated from each analysis were compared. Parent selection was also simulated based on BVs estimated from each analysis, to ascertain how differently the top parents were being ranked relative to an assumed 'true' BV (i.e. the most complete data set available of 10 years and seven additional generations of pedigree information).

This information was used to determine the minimum number of years of data and depth of pedigree needed to reliably assess parental BV in the BSES breeding program.

3.1.3. Results

Some differences in additive and additive-by-environment genetic variance estimates were observed for each data set when the six different pedigree structures were applied (Figures 1 and 2). Initial estimates of additive variance for CCS were low, but increased as additional years of data were included in analyses (Figure 1). Additive variance estimates were very similar when 3 or more years of data were included in analyses.
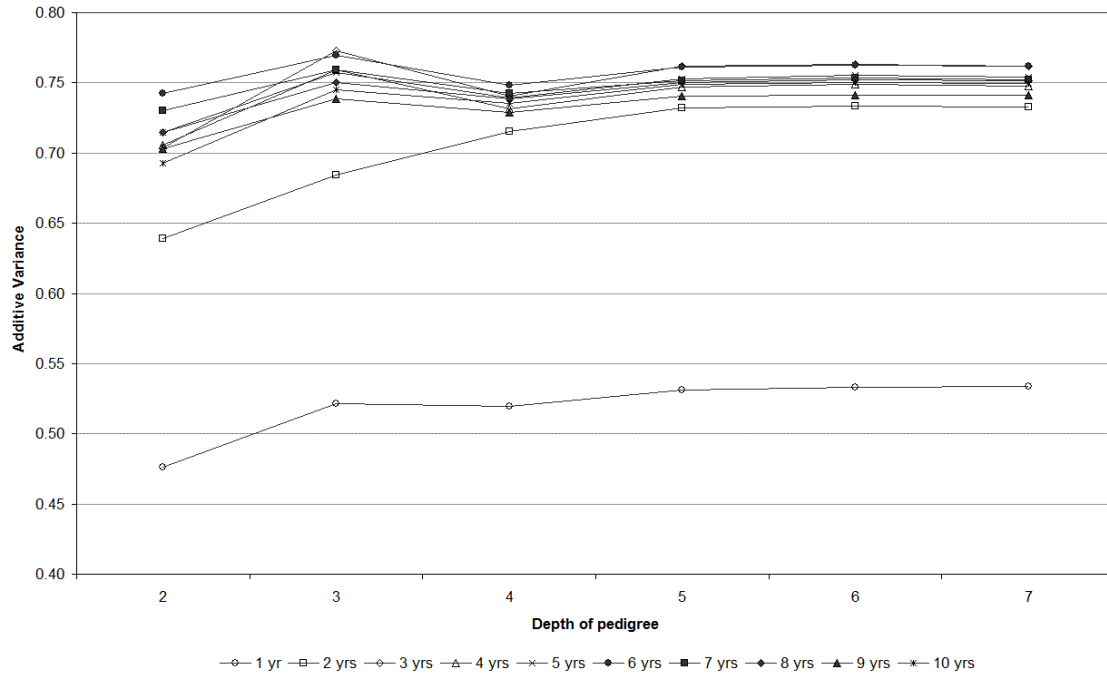
Figure 1 - Additive genetic variance estimates for each data set using different depths of pedigree for CCS for 10 years of PATs using the parental model

Additive-by-environment variance estimates showed a reverse trend to the estimates of additive variance (Figure 2), in that variance estimates decreased as the years of data increased. However, both additive and additive-by-environment estimates for each data set stabilised with the inclusion of five or more generations of pedigree information.
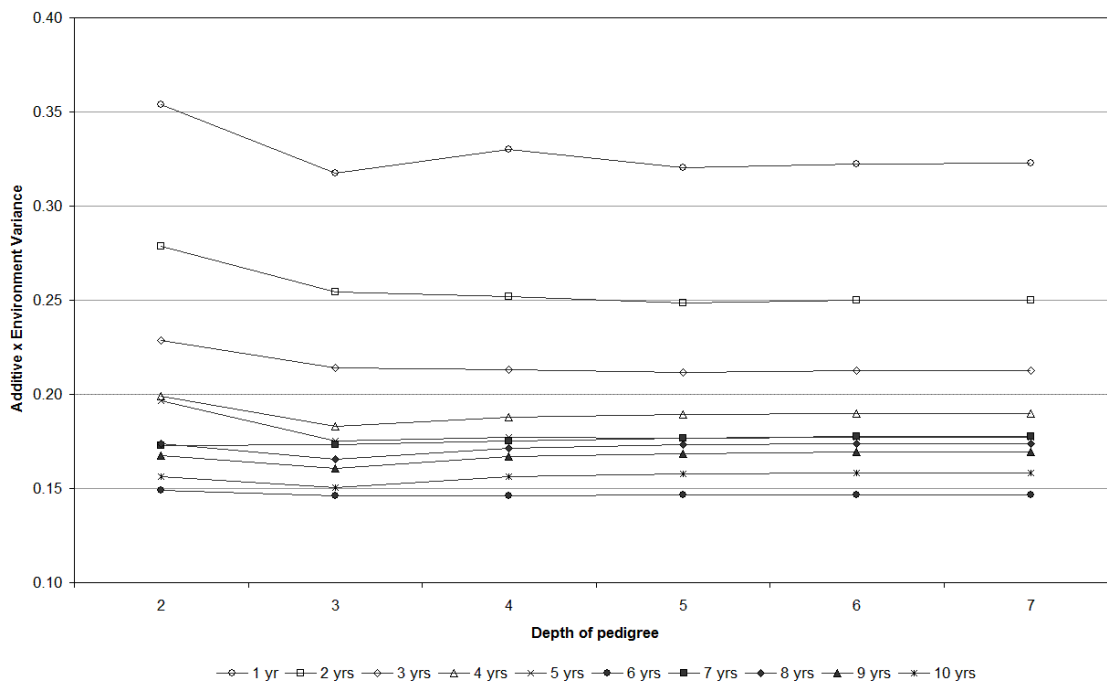
Figure 2 - Additive-by-environment variance estimates for each data set using different depths of pedigree for CCS for 10 years of PATs using the parental model

It was assumed that 'true' BV estimates for all 281 parents in common were obtained from the analysis of seven generations of pedigree and 10 years of data. There was a greatest difference between initial BV estimates using limited information (i.e. 1 year and two generations) compared with the 'true' BV (Table 1), but as years of data and/or pedigree information were added, BV estimates improved for both models when compared with 'true' BVs. Very few differences were observed in estimated BVs using at least 5 years of data and three generations of pedigree information (Table 1).

Additionally, when parents in the top 25% for BV from each analysis were identified and compared with the top parents for 'true' BV, the differences were greatest when the least amount of information was included in analyses (Table 2). As more data and pedigree information were included in analyses, the correlation coefficients for BV and number of parents in common in the top 25% increased. If only the number of years included was increased, and not depth of pedigree, the precision of estimating the BV of parents was 0.87 with a maximum of 59 of the parents in common (84%) with the top 70 parents for 'true' BV (Table 2). In addition, increasing only the depth of pedigree, and not years of data, resulted in a lower correlation coefficient and fewer parents in common compared with 'true' BV (Table 2). However, when depth of pedigree and years of data were both increased, the number of parents in common also increased considerably. When 7 or more years of data and four generations were included, at least 90% of the top parents were being identified with a correlation coefficient of 0.96 for BVs (Table 2).

Table 1 - Correlation coefficients of BVs for all 281 parents in common, comparing BV estimates using varying years of data and pedigree information with the 'true' estimate from 7 generations in the pedigree and 10 years of data for sucrose content (CCS) for the parental model

	Data sets (years of data)									
	1	2	3	4	5	6	7	8	9	10
Depth of pedigree (number of generations)										
2	0.73	0.80	0.83	0.84	0.89	0.90	0.91	0.92	0.93	0.93
3	0.77	0.84	0.90	0.91	0.96	0.97	0.98	0.99	0.99	0.99
4	0.79	0.87	0.91	0.93	0.97	0.98	0.99	0.99	1.00	1.00
5	0.80	0.88	0.91	0.93	0.97	0.98	0.99	0.99	1.00	1.00
6	0.80	0.87	0.91	0.93	0.97	0.98	0.99	0.99	1.00	1.00
7	0.80	0.87	0.91	0.93	0.97	0.98	0.99	0.99	1.00	-

Table 2 - Correlation coefficients of BVs for parents in the top 25% (i.e. the top 70 parents), comparing BV estimates using varying years of data and pedigree information with the 'true' estimate from 7 generations in the pedigree and 10 years of data for sucrose content (CCS) for the parental model

	Data sets (years of data)									
	1	2	3	4	5	6	7	8	9	10
Depth of pedigree (number of generations)										
2	0.58 (44)[†]	0.68 (49)	0.72 (51)	0.73 (52)	0.72 (54)	0.77 (55)	0.79 (58)	0.82 (59)	0.84 (58)	0.87 (59)
3	0.59 (46)	0.71 (53)	0.79 (53)	0.80 (55)	0.84 (58)	0.91 (61)	0.93 (62)	0.95 (62)	0.95 (64)	0.99 (64)
4	0.59 (48)	0.74 (55)	0.78 (55)	0.81 (56)	0.84 (61)	0.93 (62)	0.96 (65)	0.97 (65)	0.98 (67)	1.00 (68)
5	0.60 (49)	0.74 (55)	0.78 (56)	0.81 (57)	0.84 (60)	0.93 (62)	0.97 (65)	0.97 (65)	0.98 (68)	1.00 (70)
6	0.59 (48)	0.74 (56)	0.78 (56)	0.81 (57)	0.84 (60)	0.93 (63)	0.97 (65)	0.97 (66)	0.98 (68)	1.00 (70)
7	0.59 (48)	0.74 (56)	0.78 (56)	0.81 (57)	0.84 (60)	0.93 (63)	0.97 (65)	0.97 (66)	0.98 (68)	-

[†]Italic type face in parentheses shows the number of parents in common in the top 25% compared with 'true' BV estimates using 10 years of data and seven generations of pedigree.

3.1.4. Discussion

While there has been a considerable amount of research into the effects of either omitting pedigree information (Durel et al. 1998; Mehrabani-Yeganeh et al. 1999; Piepho et al. 2008) or including incorrect pedigree information (Ericsson 1999; Long et al. 1990; Visscher et al. 2002) on estimates of BV, to date, there is no known published research into the effects of different depths of pedigree on estimates of BV. In this experiment, adding pedigree information back to the base population improved the estimates of both additive variance of the population and BVs for each parent. It was also shown that minimal pedigree information yielded biased estimates of additive variance and BV, and selection accuracy of parents is low compared with using all pedigree information. Estimates of both additive variance and BV improved as pedigree information was added.

Even though it is believed that not including complete pedigree information in analyses can lead to biased estimates of additive variance and BVs (Piepho et al. 2008), in this experiment additive genetic variances were estimated with the same precision when five or more generations were included (Figure 1).

Comparisons of the accuracy to estimate BVs of sugarcane parents, including identify top-performing parents, from different amounts of data and pedigree information to the 'true' BV of each parent shows that greater amounts of information are required. If sugarcane breeders are only interested in the BVs of all the current parents, and not selecting the top parents for cross pollination, then a minimum of 5 years of data and three generations of pedigree appears to be adequate to ensure the precision of estimating BVs is very high. If the same amount of data is used to select the top parents for cross pollination, then the precision of identifying these top parents is reduced, with only 58 of the 70 top parents in common. This means that 17% of the top parents would be overlooked for cross pollination. To ensure that no more than 10% of the top parents are identified incorrectly, a minimum of 7 years of data and four generations of pedigree information is required.

As I intend to combine information from family trials from all four selection programs to more accurately predict BVs of parents in the breeding population, it is important to consider the issue of possible computing limitations in estimating BVs of over 3,200 sugarcane parents. Even though analyses are much easier and faster to process when minimal information is used, this leads to a reduction in the accuracy of estimating both additive genetic variance and BVs of parents. While including all available information (10 years of data and seven generations of pedigree) to estimate BVs for 1,600 parents in only one of the four BSES selection programs is computationally achievable, less family data and/or pedigree information will be necessary when the four selection programs are combined. It is shown that adding years of data clearly has a larger effect on the estimation of both additive variance and BVs of parents, than adding pedigree information. However, the number of years of data included in analyses should be no less than 5 years when assessing the BVs of all parents, or less than 7 years when selecting the top parents for future breeding, while the pedigree information can be reduced to three or four generations.

3.2. Breeding value-by-region interaction

3.2.1. Methods and materials

Sugarcane family trials from the four regional programs in the BSES-CSIRO Plant Improvement program in Queensland, Australia, were used for TCH and CCS from seven years of PATs established from 2001 to 2007. Sugarcane families grown each year were bi-parental cross combinations of either proven (i.e. families that have performed well in previous trials) and experimental (i.e. new or untested) combinations. A total of 5,318 unique sugarcane families were grown over the seven years across the four regional programs, with between 115 to 395 families grown each year in each region. The number of families in common across regions varied from 10% to 38% because of a highly unbalanced mating design from sparse and unreliable flowering and poor pollen fertility in sugarcane breeding (Berding and Skinner 1987). However, the connectivity of parents across the regions was greater (37% to 67%).

Each year trials were planted at the same location within each region. Growing conditions differed in average temperatures, rainfall and irrigation regimes, and soil type at each location (Table 3). Within each region, one or two trials were planted annually; and 44 trials across were planted in the 7 years for all four selection programs.

Table 3 - Weather statistics, irrigation regimes and soil types specific to each trial location on the four BSES experimental stations

Regional Selection Program	Northern	Burdekin	Central	Southern
Trial Location	Meringa	Ayr	Mackay	Bundaberg
Latitude	17°04'S	19°34'S	21°10'S	24°45'S
Yearly Average Max Temp (°C)[†]	28.8	29.6	27.5	26.6
Yearly Average Min Temp (°C)[†]	20.7	18.7	16.8	16.2
Yearly Average Rainfall (mm)[†]	1,986	1,070	1,679	1,117
Average Number of Days Rain[†]	124	63	126	97
Irrigation Type[‡]	LS[¶]	T[A]	LS	ES[B]
Location Soil Type[§]	Clifton (white schist)	Alluvial (Rudosol)	Non Calcic-Brown (Chromosol)	Red volcanic; Krasnozem (Ferresol)

[†] Meteorological data for each location sourced from Kingston (2000); [‡] irrigation information sourced from Ham et al. (2000); [§] soil type classifications; [¶] LS, limited supplementary irrigation (i.e. irrigation of pre-germinated seedlings at time of planting and continued irrigation until plants are established, once established the crop relies on natural rainfall); [A] T, total irrigation (i.e. seedling crop is totally dependent on irrigation to provide its water requirements throughout the growing period); [B] ES, extensive supplementary irrigation (i.e. irrigation is used to establish the seedling crop, and supplementary irrigation is used extensively throughout the growing period to increase water availability during months of insufficient rainfall).

Experimental data available for each of the 44 trials at time of harvest included CCS and TCH of each family plot, and were combined to assess GxE interactions for BVs of parents within and across regions.

Parent selection was also simulated to determine how differently the top parents were ranked within each region, and therefore, what impact GxE interactions present across regions have on parental BVs estimated.

3.2.2. Statistical analysis

A mixed model was used to partition genotypic effects for each parent into additive and non-additive genetic effects using the parental model while including site-specific spatial variation in each trial.

The progression of analyses performed for assessing both within and across region GxE for TCH and CCS is outlined in Table 4. The variance component model was used to estimate the size of additive and additive by environment variance components, and whether more sophisticated models were needed. It was also used as starting values for the Factor-Analytic (FA) model.

Table 4 - Models used to assess genotype by environment interactions both within and across the four regional selection programs in the BSES breeding program

Model sequence	Random genetic (G) term		Residual error (R) term	
	Parent.Trial/Region		column.row	
	Model	Comments	Model	Comments
1	Variance component model (VC)	No genetic (G) structure (homogeneous variance for all environments)	Spatial model	Residual error (R) structure (two-dimensional separable autoregressive spatial model for error)
2	Uniform correlation model (CORR)	G structure (heterogeneous variance and homogeneous covariance between environments)	Spatial model	(as above)
3	Factor-Analytic model (FA)	G structure (heterogeneous variance and covariance between environments)	Spatial model	(as above)

3.2.3. Results

Genetic correlations within each region (not presented here) ranged between 0.79 and 0.96 for TCH, and 0.71 and 0.84 for CCS. Similar correlation coefficients were found between pairs of regions for both traits (Table 5); however, genetic correlations between the Burdekin and any other region were lower for both TCH and CCS.

Table 5 - Variance/correlation matrix from the Factor-Analytic analysis of sugarcane family data for estimating breeding values (BVs) of sugarcane parents for cane yield (TCH) and sucrose content (CCS). Numbers on the lower diagonal in bold type face are the additive variance components estimated for each regional selection program for TCH; numbers on the upper diagonal in italic bold type face are the additive variance components estimated for each regional selection program for CCS; numbers in the lower off-diagonal are the genetic correlations between pairs of regions for TCH; and the numbers in italic type face in the upper off-diagonal are the genetic correlations between pairs of regions for CCS.

			Northern	*Burdekin*	*Central*	*Southern*	*Region*	
				0.353	*0.57*	*0.78*	*0.82*	*Northern*
		Northern	**18.13**		***0.230***	*0.64*	*0.68*	*Burdekin*
TCH		Burdekin	0.79	**118.3**		***0.131***	*0.93*	*Central*	*CCS*
		Central	0.95	0.79	**43.76**		***0.203***	*Southern*
		Southern	0.83	0.68	0.82	**115.9**		
	Region		Northern	Burdekin	Central	Southern		

The proportion of top parents in common when parent selection was simulated varied between pairs of regions (Table 6). The regions with the most parents in common were the Northern and Central regions for TCH, the Central and Southern regions for CCS. Similar to the genetic correlations estimated, the Burdekin region had the fewest parents in common with all other regions.

Table 6 - Percentage of common parents selected for breeding values between pairs of regions for cane yield (TCH) and sucrose content (CCS) estimated from the Factor-Analytic analysis of sugarcane family data. Numbers on the upper off-diagonal in italic type face are percentage (%) of parents in common in the top 25% between pairs of regions for CCS; numbers in the lower off-diagonal are percentage (%) of parents in common in the top 25% between pairs of regions for TCH.

		Northern	*Burdekin*	*Central*	*Southern*	*Region*	
TCH	Northern		*55*	*67*	*78*	*Northern*	CCS
	Burdekin	78		*64*	*55*	*Burdekin*	
	Central	89	69		*81*	*Central*	
	Southern	68	58	66		*Southern*	
Region		Northern	Burdekin	Central	Southern		

3.2.4. Discussion

Currently, BSES maintains separate parent populations for each regional breeding program because of perceived GxE interactions that exist between growing regions which allegedly give different responses among sugarcane parent performance. This has led to plant breeders selecting sugarcane parents for cross pollination that are adapted to each of the four distinct regions. However, it is debatable wether this is the best way to manage the BSES-CSIRO Plant Improvement program and the needs of the industry, or whether BSES could improve the process of selecting parents across the sugarcane regions as a whole. Therefore, it was important to establish the presence or absence of GxE interactions across the industry, and understand how GxE affects estimating BVs for each region, and parent selection for cross pollination.

In each regional breeding program, sugarcane families have always been grown at the same location each year. Therefore, it was not surprising to find that the magnitude of GxE interactions within each region was small compared to genotypic effects, and resulted in relatively high genetic correlations between trials within each region for TCH and CCS.

Similar genetic correlations were estimated between pairs of regions for TCH and CCS indicating that BVs were more stable across the four regions. The regions which had the greatest associations for both TCH and CCS were the Central and Northern regions, and the Central and Southern regions (Table 5). Both the Central and Northern regions have limited supplementary irrigation, while the Central and Southern regions are more similar geographically, and hence, climatically. These common environmental conditions in which sugarcane families were grown might explain the high genetic correlations and similar parent selections for both TCH and CCS, indicating that parents' performance responses were similar across these regions. In contrast, of the four regions the Burdekin had the lowest genetic correlations with all other regions. This was reflected in there being only just over half of the top parents selected in the Burdekin in common with other regions (Table 6). This may be because the sugarcane family trials were grown under very different conditions in the Burdekin region compared to other regions. For example, average cane yields for family trials in the Burdekin were between 20 and 45 tonnes cane per hectare greater than the other regions because sugarcane families were grown under high-input irrigated conditions.

Despite there being some differences in parents selected in each regional program, BVs from a given region for TCH and/or CCS were of great relevance in predicting specific parental performance in any other region. The genetic correlations indicate that there are associations between additive genetic merits between pairs of regions for TCH and CCS, and similar correlations were seen between regions for unselected sugarcane clones (Chapman et al. 2004). In addition to the good correlations between GxE interactions within and across regions, the fact that the magnitude of GxE within and across regions was of similar size suggests that the important factors driving GxE in family trials are sampled within any individual region rather than being region specific.

The results from this experiment should lead to better co-ordination of family trials, parent selection, and use of data and varietal exchange between different regions. For example, in the cases where parent performance were similar, such as the Central and Southern regions, there may be opportunities for such breeding programs to benefit from combining parent populations for cross pollination and/or family trials to test cross combinations. Joint analysis of all family trials means it is also feasible to indirectly select parents for cross pollination using data from other regions.

3.3. Comparison of the new statistical approach developed in this thesis to the previous approach for sugarcane parent selection

3.3.1. Methods and Materials

Ten years of sugarcane family trials or PATs from 1998 to 2007 from the four regional programs in the BSES-CSIRO Plant Improvement program in Queensland, Australia, for TCH and CCS content were used. The families included in trials were bi-parental crosses of either proven (i.e. families that have performed well in previous PATs) or experimental (i.e. new or untested) combinations. An average of 273 unique families were grown annually per region, and a total of 6,954 families were grown in 67 separate trials planted over ten years across the four regions.

Data from the 67 individual trials were combined to generate five data sets. As regions were analysed independently using the previous statistical approach, four data sets were created comprising of ten years of data from each region for individual analysis. Trial data from the seven most recent years (Atkin et al. 2009a) from all regions were combined for the new statistical approach. Pedigree files were also generated for each data set. The four pedigree files generated for the individual regions contained all pedigree information available for each family back to their founding parents, while the pedigree file generated for the new approach where four regions were combined contained only four generations of pedigree information as recommended by Atkin et al. (2009a).

Analyses were conducted for TCH and CCS using both the new and previous approaches to estimate BVs. BVs were compared between approaches using simulated selection of sugarcane parents without using a selection index. BVs were then incorporated into two selection indexes, Net Merit Grade (NMG) for the previous approach and Economic Breeding Value ($EBV) for the new approach. The accuracy of predicting BVs using both approaches was also calculated and compared by estimating confidence ranges for each BV for each trait using Bourdon's (2000) formula.

3.3.2. Statistical Analysis

Differences in the statistical approach taken to estimate BVs using the previous and new approach developed in this thesis are given in Table 7.

Table 7 - Key differences between the previous and the new approach of estimating breeding values of sugarcane parents in the BSES-CSIRO Plant Improvement program

Analytical Information	Previous Approach	New Approach
BLUP analysis on trait/s	Net Merit Grade (selection index derived from cane yield & sucrose content)	Cane yield & sucrose content analysed independently
Data set used	10 years x 1 regional program	7 years x 4 regional programs
Amount of pedigree information used	All (back to founding parents)	4 generations of pedigree
Genetic model used	Bi-parental model	Bi-parental model
G structure used	No (only used a simple variance component model)	Yes (Factor-Analytic model)
Non-additive genetic effects partitioned	No	Yes
Interactions estimated	No (only main additive effect estimated)	Yes (genetic correlations between regions for a given trait estimated)
Site-specific spatial variation accounted for	No	Yes
Analysis of regional programs	Regional programs analysed independently (but breeding values averaged across regions)	Combined analysis of all regional programs
Selection index used for parental selection	Parents ranked and selected based on breeding values estimated for Net Merit Grade only	Parents ranked and selected based on Economic Breeding Values ($EBVs) generated from combined breeding values estimates for each trait and their associated economic weightings using the new Genetic Evaluation System (GES)

3.3.3. Results

There was a 29% increase in expected genetic gain, equivalent to additional gains of 1.5 tonnes of cane per hectare, when parents were selected for TCH using the new approach over the previous approach (Table 8) across all regional programs. Selecting parents for CCS using the new approach also resulted in additional gains of 0.06% sucrose content (a 10% increase) (Table 8). In addition, selection of parents for TCH and CCS simultaneously using the new approach resulted in a gain or profit increase of 12% from producing one tonne of sugar (Table 8).

Table 8 - Additional genetic gains for cane yield (TCH) and sucrose content (CCS) from selection of parents using the proposed new statistical approach over the previous approach assuming independent selection of traits, plus the additional gains/profits in dollars per tonne of sugar produced from selection of parents using the proposed new statistical approach over the previous approach assuming simultaneous selection of traits

Trait/s	Units of gain	Region	Mean BV/$EBV ± SE		Additional gain	
			New approach	Previous approach	Unit gain	Percentage gain (%)
TCH	Tonnes of cane per hectare	Northern	3.84[†] ± 0.14	3.03[‡] ± 0.20	0.81	26.7
		Burdekin	7.43[†] ± 0.39	4.95[‡] ± 0.55	2.48	48.3
		Central	5.98[†] ± 0.22	4.98[‡] ± 0.29	1.00	20.1
		Southern	9.34[†] ± 0.43	7.65[‡] ± 0.54	1.69	22.1
		All Regions	6.61[†] ± 0.23	5.11[‡] ± 0.34	1.50	29.4
CCS	% sucrose content	Northern	0.87[†] ± 0.02	0.79[‡] ± 0.03	0.08	10.1
		Burdekin	0.57[†] ± 0.02	0.54[‡] ± 0.02	0.03	5.6
		Central	0.55[†] ± 0.01	0.50[‡] ± 0.02	0.05	10.0
		Southern	0.73[†] ± 0.02	0.66[‡] ± 0.02	0.07	10.6
		All Regions	0.68[†] ± 0.01	0.62[‡] ± 0.02	0.06	9.7
$EBV (TCH & CCS)	$ per tonne of sugar produced	Northern	13.04[§] ± 0.40	11.68[¶] ± 0.50	1.36	11.6
		Burdekin	9.06[§] ± 0.37	7.24[¶] ± 0.52	1.82	25.2
		Central	13.34[§] ± 0.37	12.34[¶] ± 0.47	1.00	8.1
		Southern	20.88[§] ± 0.62	19.20[¶] ± 0.77	1.68	8.8
		All Regions	14.03[§] ± 0.28	12.57[¶] ± 0.44	1.46	11.6

[†] mean BV of parents selected using the proposed new statistical approach for a given trait; [‡] mean BV of parents selected using the previous statistical approach for a given trait; [§] mean $EBV of parents selected for TCH and CCS simultaneously using the proposed new statistical approach; [¶] mean $EBV of parents selected for TCH and CCS simultaneously using the previous statistical approach.

BVs predicted using the previous and proposed new statistical approaches were used to generate NMG and $EBV, respectively. When the two different selection indices were compared, parents were re-ranked when I based selections on $EBVs generated using the improved BV estimates instead of NMG, with fewer than 60% of the top parents in common when selection was based on both indices (Figure 3).

$EBV versus NMG
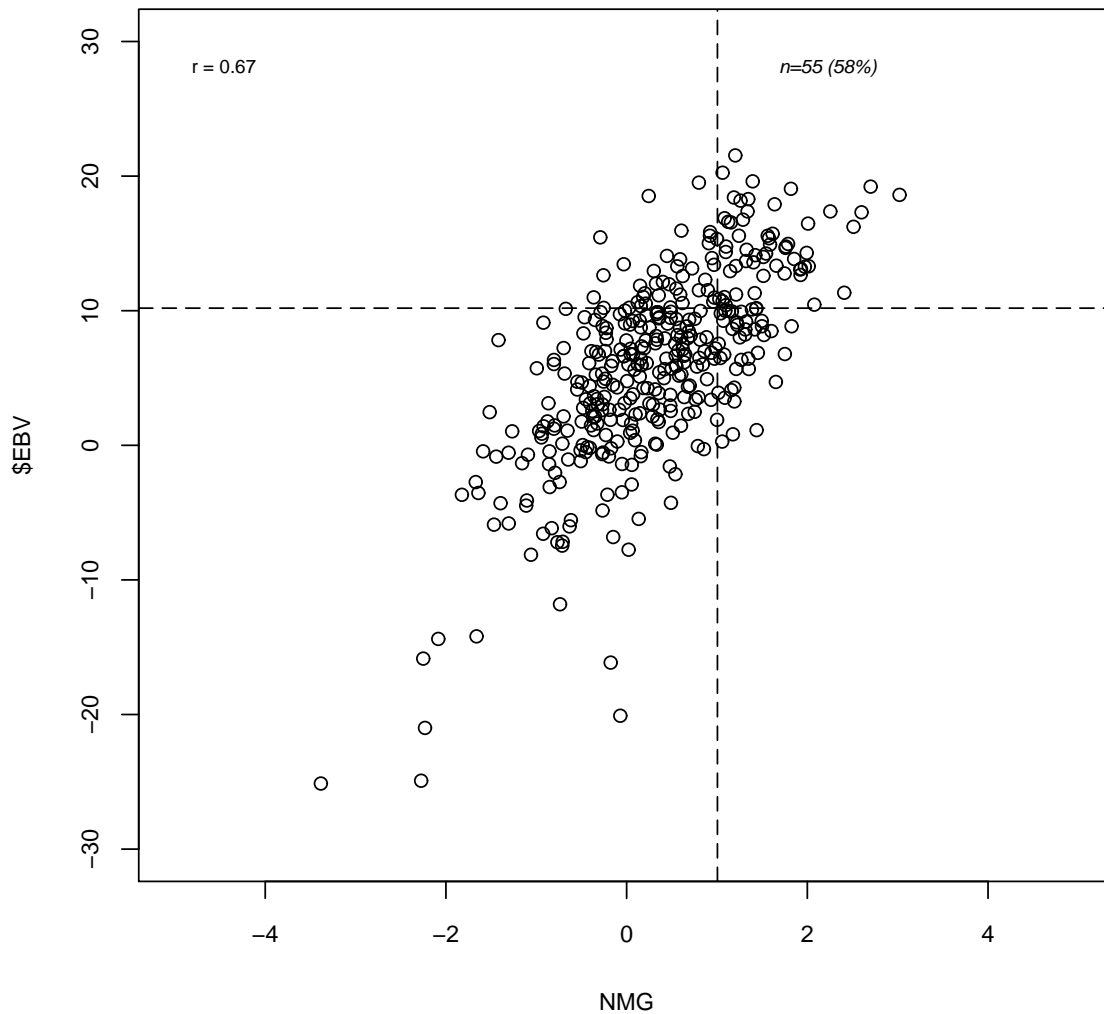
Figure 3 - Comparison of the previous selection index using Net Merit Grade (NMG) and the new selection index using Economic Breeding Value ($EBV). Black dashed lines represent the threshold where the top parents would be selected for using the previous statistical approach and selection index NMG (vertical) and selected for using the proposed new statistical approach and selection index $EBV. In the upper right quadrant are the top parents in common using both statistical and selection index approaches, in the lower right quadrant are parents selected for NMG only, and in the upper left quadrant are parents selected for $EBV only. Correlation coefficient between BVs and selection indices and the number of parents in common from selection using the two approaches are also given.

Differences in additional gains were observed when I selected parents for $EBV over NMG for the two traits TCH and CCS (Table 9), in contrast to gains observed when I compared BVs for both approaches without using a selection index. Parents were selected for $EBV over NMG reduced their expected genetic gain by of 0.5 tonnes of cane per hectare (11%) for TCH across the four regional program, while an additional gain of 0.17 percent sucrose content (47%) for CCS was estimated. However, there was a 20% gain in profit from producing one tonne of sugar per generation (or $2.34/ts) when top parents were selected for $EBV over NMG (Table 9).

Gains for traits within each regional program were more relatively similar when parents were selected using a selection index (Table 9) in comparison to when parents were selected using BVs of individual traits (Table 8). However, trends in expected profit gains for $/ts produced from parent selection were different across the regions (Table 9). Gains in the Northern region were the highest at 33% of the four regions and were twice the size of gains in the Central and Southern regions which only had approximately 15% profit gains from producing one tonne of sugar.

Table 9 - Expected genetic gains (dollars per tonne of sugar) from selection of top parents (top 25%) by Economic Breeding Value ($EBV) and Net Merit Grade (NMG) in four regions combined based on Progeny Assessment Trials. The impact or expected genetic gain of individual traits for cane yield (TCH) and sucrose content (CCS) are also given

Trait/s	Units of gain	Region	Mean BV/$EBV ± SE		Additional gain	
			$EBV	NMG	Unit gain	Percentage gain (%)
TCH	Tonnes of cane per hectare	Northern	2.67 ± 0.21	3.04 ± 0.19	-0.37	-12.2
		Burdekin	4.09 ± 0.55	4.59 ± 0.53	-0.50	-10.9
		Central	4.18 ± 0.34	4.75 ± 0.29	-0.57	-12.0
		Southern	6.64 ± 0.52	7.27 ± 0.53	-0.63	-8.7
		All Regions	4.39 ± 0.36	4.91 ± 0.33	-0.52	-10.6
CCS	% sucrose content	Northern	0.72 ± 0.03	0.48 ± 0.04	0.24	50.0
		Burdekin	0.44 ± 0.02	0.30 ± 0.03	0.14	46.7
		Central	0.46 ± 0.02	0.33 ± 0.03	0.13	39.4
		Southern	0.60 ± 0.02	0.42 ± 0.03	0.18	42.9
		All Regions	0.56 ± 0.02	0.38 ± 0.03	0.18	47.4
$EBV/NMG[†]	$ per tonne of sugar produced	Northern	13.20 ± 0.41	9.91 ± 0.61	3.29	33.2
		Burdekin	8.74 ± 0.40	7.23 ± 0.48	1.51	20.9
		Central	13.30 ± 0.36	11.56 ± 0.51	1.74	15.1
		Southern	20.87 ± 0.62	18.05 ± 0.78	2.82	15.6
		All Regions	14.03 ± 0.28	11.69 ± 0.48	2.34	20.0

[†] BVs for TCH and CCS estimated independently for $EBV using the proposed new statistical approach, and BVs for NMG estimated using the previous statistical approach.

Confidence ranges estimated for BV predictions for TCH and CCS reduced when the new statistical approach was used over the previous approach (Table 10). Reductions in confidence intervals of BV predictions ranged from 30% to 65% for TCH and 23% to 55% for CCS. Average reductions in confidence intervals across all regions were 48% and 37% for TCH and CCS respectively.

Table 10 - Confidence ranges of breeding values predicted from two different statistical approaches for each BSES regional program for the top parents

Trait	Region	New approach		Previous approach	
		Mean BV ± CI[†]	95% Confidence range	Mean BV ± CI	95% Confidence range
Cane yield (tonnes of cane per hectare)	Northern	2.67 ± 3.11	-0.44 to 5.78	3.04 ± 8.80	-5.76 to 11.84
	Burdekin	4.09 ± 8.13	-4.04 to 12.22	4.59 ± 12.00	-7.41 to 16.59
	Central	4.18 ± 4.68	-0.50 to 8.86	4.75 ± 10.19	-5.44 to 14.94
	Southern	6.64 ± 7.27	-0.63 to 13.91	7.27 ± 10.33	-3.06 to 17.60
	All regions	4.39 ± 5.39	-1.00 to 9.78	4.91 ± 10.27	-5.36 to 15.18
Sucrose content (% sucrose content)	Northern	0.72 ± 0.39	0.33 to 1.11	0.48 ± 0.51	-0.03 to 0.99
	Burdekin	0.44 ± 0.35	0.05 to 0.79	0.30 ± 0.57	-0.27 to 0.87
	Central	0.46 ± 0.23	0.23 to 0.69	0.33 ± 0.51	-0.18 to 0.84
	Southern	0.60 ± 0.27	0.33 to 0.87	0.42 ± 0.35	0.07 to 0.77
	All regions	0.56 ± 0.31	0.25 to 0.87	0.36 ± 0.49	-0.13 to 0.85

[†] CI, confidence interval.

3.3.4. Discussion

A new statistical approach was developed to improve the ability of BSES sugarcane breeders to identify and select elite parents for future cross pollination in the BSES-CSIRO Plant Improvement program. By using the new statistical approach outlined here, I were able to use sugarcane family data more efficiently (Atkin et al. 2009a) and combine data across the four regional programs to estimate the BVs for sugarcane clones for parental selection with improved accuracy. This is because the new approach uses more information on sugarcane families and parents as many parents are common across the regions, and an analysis that combines trials predicts genetic effects more accurately (Bourdon 2000; Hammond et al. 1992).

In addition, even though parents selected for NMG using the previous statistical approach had higher BVs for TCH on average than parents selected for $EBV using the new statistical approach, note that the lower limit to the confidence range is considerably less than zero (Table 10). Therefore, breeders cannot be as sure that the top parents selected using the old approach are better than the average of the population. However, in using the new approach I can be reasonably confident that the parents selected for future cross pollination are better than the population average.

The improved accuracy of predicting BVs was also due to other improvements made to the previous statistical approach including:

- reducing the potential of producing biased estimates of BVs by partitioning non-additive genetic effects (Henderson 1985; Quinton and Smith 1997; Wei and Van der Werf 1993);
- accounting for spatial variation present in many of the family trials by partitioning spatial variation (Stringer 2006; Stringer and Cullis 2002), and;
- using the Factor-Analytic model by accommodating and exploiting genetic correlations across the four regions to further improve estimates of BVs.

Parents were re-ranked using BV estimates from the new approach resulting in differences in parental selection when the two approaches were compared. This improved efficiency in identifying top parents led to additional increases in expected genetic gains of approximately 29%, 10% and 12% in tonnes of cane per hectare, percentage sucrose content and profits for one tonne of sugar produced respectively per generation by using the new statistical approach to estimate BVs over the previous approach.

A further 8% increase in overall genetic gain with additional profits of $2.34 from producing one tonne of sugar was achieved by selecting these high BV parents over parents selected using the previous approach and NMG. Further additional gains or profits could be achieved through parent selection once more traits such as disease information is included in generating $EBV.

4.0   OUTPUTS

The major output from this project is a more efficient approach to estimating the breeding potential of current sugarcane parents for BSES plant breeders. Breeders can now be more confident in their parent selection for cross pollination, and selection of specific cross combinations using more accurate estimates of parental breeding values. This has been implemented within the BSES-CSIRO plant improvement program. Other outputs from this project include international scientific papers and a PhD thesis.

5.0   OUTCOMES

This project has resulted in an improved method of estimating breeding values of sugarcane parents. This is likely to result in increased genetic gain through more accurate assessment and selection of sugarcane parents and specific cross combinations through cross pollination; hence, the production and advancement of superior sugarcane clones through the various stages of selection, and finally the release of elite commercial varieties to the Australian sugarcane industry.

6.0   FUTURE NEEDS AND RECOMMENDATIONS

I have demonstrated that interplot competition is present in family trials (Atkin et al. 2009b) and affects estimates of TCH of family plots; however, it has proved difficult applying the competition model to larger data sets, such as combining all family trials across regional programs. Therefore, further research is needed in consultation with NSW DPI to investigate using the interplot competition model in sugarcane family trials.

As sugarcane parental BVs can also be estimated from later stage trials, Stages 2 and 3, BVs estimated from family trials (Stage 1) and Stages 2 and 3 combined, and the accuracy of these estimates, should be compared. Therefore, BSES breeders can apply appropriate weightings to different BVs estimated from various sources.

7.0   PUBLICATIONS ARISING FROM THE PROJECT

Atkin FC, Dieters MJ, Stringer JK (2009a) Impact of depth of pedigree and inclusion of historical data on the estimation of additive variance and breeding values in a sugarcane breeding program. Theoret Appl Genet 119:555-565 (Appendix 2)

Atkin FC, Stringer JK, Dieters MJ (2009b) Interplot competition between sugarcane families significantly affects estimates of cane yield and family selection. Proc Aust Soc Sugar Cane Technol 31:326-331 (Appendix 3)

A manuscript on a comparison of the new statistical approach developed in this thesis and the previous approach had been submitted to Theoretical and Applied Genetics (TAG) journal (see Appendix 4) and I am currently waiting for comments from reviewers. Another manuscript on GxE interactions between parental breeding values in the Australian sugarcane industry is also being prepared for submission to the same journal. These are based mainly on Chapter 7 and 6 respectively in the PhD thesis.

8.0   ACKNOWLEDGEMENTS

9.0   REFERENCES

Atkin FC, Dieters MJ, Stringer JK (2009a) Impact of depth of pedigree and inclusion of historical data on the estimation of additive variance and breeding values in a sugarcane breeding program. Theoret Appl Genet 119:555-565

Atkin FC, Stringer JK, Dieters MJ (2009b) Interplot competition between sugarcane families significantly affects estimates of cane yield and family selection. Proc Aust Soc Sugar Cane Technol 31:326-331

Berding N, Skinner JC (1987) Traditional breeding methods. Copersucar International Sugarcane Workshop, Copersucar, Piracicaba, Brazil

Bourdon RM (2000) Understanding animal breeding. Prentice-Hall Inc., New Jersey

Chapman SC, Jackson PA, Rattey A (2004) Genotype by environment interactions across the Australian sugarcane industry. Proc Aust Soc Sugar Cane Technol 26:8 p

Costa e Silva J, Borralho NMG, Potts BM (2004) Additive and non-additive genetic parameters from clonally replicated and seedling progenies of *Eucalyptus globulus*. Theoret Appl Genet 108:1113-1119

Durel CE, Laurens F, Fouillet A, Lespinasse Y (1998) Utilization of pedigree information to estimate genetic parameters from large unbalanced data sets in apple. Theoret Appl Genet 96:1077-1085

Ericsson T (1999) The effect of pedigree error by misidentification of individual trees on genetic evaluation of a full-sib experiment. Silvae Genet 48:239-242

Gilmour AR, Cullis BR, Verbyla AP (1997) Accounting for natural and extraneous variation in the analysis of field experiments. J Agric Biol Environ Stat 2:269-293

Gilmour AR, Gogel BJ, Cullis BR, Thompson R (2006) ASReml User Guide. 2.0 edn. VSN International Ltd, Hemel Hempstead, UK

Ham G, McGuire P, Kingston G (2000) Irrigation of sugarcane. In: Hogarth DM, Allsopp PG (eds) Manual of Canegrowing. Bureau of Sugar Experiment Stations, Brisbane, pp 195-225

Hammond K, Graser HU, McDonald A (1992) Animal breeding the modern approach: A textbook for consultants, farmers, teachers and for students of animal breeding. University of Sydney, Sydney

Henderson CR (1984) Applications of linear models in animal breeding. University of Guelph, Guelph, Canada

Henderson CR (1985) Best linear unbiased prediction of nonadditive genetic merits in noninbred populations. J Anim Sci 60:111-117

Kelly AM, Cullis BR, Gilmour AR, Eccleston JA, Thompson R (2009) Estimation in a multiplicative mixed model involving a genetic relationship matrix. Genet Select Evol 41:9 p

Kingston G (2000) Climate and the management of sugarcane. In: Hogarth M, Allsopp P (eds) Manual of Canegrowing. Bureau of Sugar Experiment Stations, Brisbane, pp 7-25

Long TE, Johnson RK, Keele JW (1990) Effects of errors in pedigree on three methods of estimating breeding value for litter size, backfat and average daily gain in swine. J Anim Sci 68:4069-4078

Mehrabani-Yeganeh H, Gibson JP, Schaeffer LR (1999) Using recent versus complete pedigree data in genetic evaluation of a closed nucleus broiler line. Poult Sci 78:937-941

Mrode RA (2005) Linear models for the prediction of animal breeding values, 2nd edn. CABI Publishing, Wallingford, UK

Piepho HP, Mohring J, Melchinger AE, Buchse A (2008) BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161:209-228

Quinton M, Smith C (1997) An empirical check on best linear unbiased prediction genetic evaluation using pig field recording data. Can J Anim Sci 77:211-216

Roach BT (1989) Origin and improvement of the genetic base of sugarcane. Proc Aust Soc Sugar Cane Technol 11:34-48

Smith A, Cullis B, Thompson R (2001) Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. Biometrics 57:1138-1147

Stefanova KT, Smith AB, Cullis BR (2009) Enhanced diagnostics for the spatial analysis of field trials. J Agric Biol Environ Stat 14:392-410

Stringer JK (2006) Joint modelling of spatial variability and interplot competition to improve the efficiency of plant improvement. The University of Queensland, Brisbane, p 224

Stringer JK, Cullis BR (2002) Application of spatial analysis techniques to adjust for fertility trends and identify interplot competition in early stage sugarcane selection trials. Aust J Agric Res 53:911-918

Visscher PM, Woolliams JA, Smith D, Williams JL (2002) Estimation of pedigree errors in the UK dairy population using microsatellite markers and the impact on selection. J Dairy Sci 85:2368-2375

Wei M, Van der Werf JHJ (1993) Animal model estimation of additive and dominance variances in egg production traits of poultry. J Anim Sci 71:57-65

APPENDIX 1 – DRAFT PhD THESIS

Electronic version only.

APPENDIX 2 – ATKIN et al. (2009a)

APPENDIX 3 – ATKIN et al. (2009b)

APPENDIX 4 – ATKIN et al. (2010)